

REDUCTION OF BACKGROUND NOISE FOR SPEECH ENHANCEMENT

Publication number: JP9503590 (T)

Publication date: 1997-04-08

Inventor(s):

Applicant(s):

Classification:

- international: G10L13/00; G10L21/02; H03H17/02; H04B1/10; H04B15/00; H04S1/00; G10L13/00; G10L21/00; H03H17/02; H04B1/10; H04B15/00; H04S1/00; (IPC1-7): G10L9/00; H03H17/02; H04B15/00

- European: G10L21/02A1; H04B1/10E; H04S1/00A

Application number: JP19940504026T 19940606

Priority number(s): WO1994US06367 19940606; US19930086707 19930707

Also published as:

WO9502288 (A1)

US5550924 (A)

JP3626492 (B2)

EP0707763 (A1)

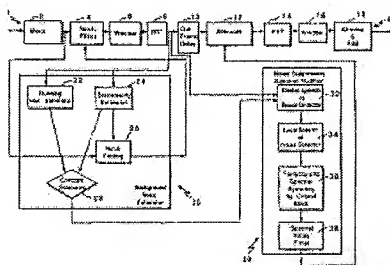
EP0707763 (A4)

more >>

Abstract not available for JP 9503590 (T)

Abstract of corresponding document: WO 9502288 (A1)

Properties of human audio perception are used to perform spectral and time masking to reduce perceived loudness of noise added to speech signals. A signal is divided into blocks (2), passed through notch filters (4) to remove noise components and then appended to part of the previous block (6). An FFT (8) is then performed on the resulting block and the spectral components are fed to noise estimator (20). Each frequency component is then analyzed to determine whether it is noise. The frequency component's gain function is determined and a spectral valley filler (38) is used to process the gain function after which the function is used to modify magnitude components of the FFT (12). In inverse FFT (14) then maps the signal back to the time domain to give a frame of noise-reduced signal.



Data supplied from the **espacenet** database — Worldwide

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表平9-503590

(43) 公表日 平成9年(1997)4月8日

(51) Int.Cl. ⁸	識別記号	序内整理番号	F I	
G 1 0 L 9/00		9381-5H	G 1 0 L 9/00	F
H 0 3 H 17/02	6 0 1	9274-5J	H 0 3 H 17/02	6 0 1 G
H 0 4 B 15/00		9298-5J	H 0 4 B 15/00	

審査請求 未請求 予備審査請求 有 (全 29 頁)

(21) 出願番号 特願平7-504026
(86) (22) 出願日 平成6年(1994)6月6日
(85) 翻訳文提出日 平成8年(1996)1月5日
(86) 国際出願番号 PCT/US94/06367
(87) 国際公開番号 WO95/02288
(87) 国際公開日 平成7年(1995)1月19日
(31) 優先権主張番号 08/086, 707
(32) 優先日 1993年7月7日
(33) 優先権主張国 米国 (US)
(81) 指定国 EP (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), CA, JP

(71) 出願人 ビクチャーテル コーポレイション
アメリカ合衆国 マサチューセッツ州
01960 ビーボディ ワン コーポレイシ
ョン ウェイ (番地なし)
(72) 発明者 ヘルプ ブラント マーティン
アメリカ合衆国 マサチューセッツ州
02176 メルローズ マウント ヴァーノ
ン ストリート 83
(72) 発明者 チュー ピーター エル
アメリカ合衆国 マサチューセッツ州
02173 レキシントン ハードリー ロー
ド 7
(74) 代理人 弁理士 中村 稔 (外6名)

(54) 【発明の名称】 会話の品質向上のための背景雑音の低減

(57) 【要約】

人間の音に対する知覚の特性が、スペクトル・マスクおよび時間マスクを行うために使用され、会話信号に付加された知覚される雑音の強度を低減する。信号は、ブロック(2)に分割され、ノッチ・フィルタ(4)を通過して、雑音成分が低減され、先行するブロック(6)の一部を追加される。続いて、FFT(8)が、結果のブロックに施され、スペクトル成分は雑音評価器(20)に与えられる。続いて、各周波数成分は分析され、雑音かどうか判断される。周波数成分の利得関数が決定され、スペクトル・パレー充填器(38)が、利得関数进行处理するために使用される。その後、この関数は、FFT(12)の大きさの成分を修正するために使用される。続いて、逆FFT(14)において、信号は、時間領域に写像され、雑音の低減された信号フレームが得られる。

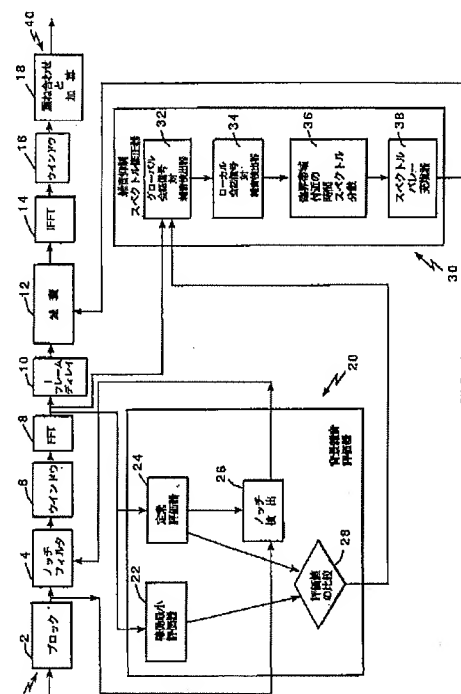


FIG. 1

【特許請求の範囲】

1. 雑音成分を有する入力音響信号に背景雑音の知覚される抑制を行うための装置であって、

前記入力音響信号を複数の信号フレームに分割するフレーム化装置、

前記信号フレームから狭帯域の雑音成分を除去して、フィルタリングされた信号フレームを生成するノッチ・フィルタ・バンク、

一つのフィルタリングされた信号フレーム内の全信号を、前記一つのフィルタリングされた信号フレームに時間的に直前に先行するフィルタリングされた信号フレーム内のいくつかの信号と結合し、ウィンドウ化された信号フレームを生成するウィンド化装置、

前記ウィンドウ化された信号フレームから周波数スペクトル成分を得る変換器、

前記周波数スペクトル成分を使用して、前記周波数スペクトル成分の雑音量の雑音評価値を生成する雑音評価器、

前記雑音評価値および前記周波数スペクトル成分に基づいて、利得乗法因数を生成する雑音抑制スペクトル修正器、

前記周波数スペクトル成分を遅延させ、遅延した周波数スペクトル成分を生成する遅延器、

前記利得乗法因数に基づいて前記遅延した周波数スペクトル成分を減衰させ、雑音の低減された周波数成分を生成する制御減衰器、および

前記雑音の低減された周波数成分を時間領域に変換する逆変換器、
を備えている装置。

2. 前記雑音抑制スペクトル修正器が、

前記周波数スペクトル成分の各周波数成分に対して、その周波数成分が雑音であるかどうかについての決定を行うグローバル判定機構、

前記周波数スペクトル成分の各周波数成分に対して、周波数成分が雑音成分である確度レベルを導出するローカル雑音判定機構、

前記確度レベルに基づいて、各周波数成分に対して前記利得乗法因数を決定

する検出器、

前記利得乗法因数をスペクトル的および時間的に調整する分散機構、ならびに

前記周波数成分のスペクトルの谷間を検出および充填するスペクトル・バレー充填器、

を備えている請求の範囲第1項に記載の装置。

3. 前記背景雑音評価器が、各周波数スペクトル成分に対して、対応する雑音評価値を生成するものであり、

前記ローカル雑音判定機構が、

(a) 周波数成分のそれぞれとその対応する雑音評価値との比、および

(b) 前記グローバル判定機構によって行われた決定、

に基づいて、確度レベルを導出するものである、

請求の範囲第2項に記載の装置。

4. 前記分散機構が、前記利得乗法因数を前記確度レベルに基づいて調整するものである請求の範囲第2項に記載の装置。

5. 滑らかにされた時間領域の成分を生成して、前記雑音の低減された時間領域の成分の不連続性を最小化する後置ウィンドウ、および

前記滑らかにされた時間領域の成分の第1の部分を、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、前記滑らかにされた時間領域の成分の前記第1の部分に含まれない部分からなる残りの部分を格納する重ね合わせ加算器、

をさらに備えている請求の範囲第1項に記載の装置。

6. 滑らかにされた時間領域の成分を生成して、前記雑音の低減された時間領域の成分の不連続性を最小化する後置ウィンドウ、および

前記滑らかにされた時間領域の成分の第1の部分を、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、前記滑らかにされた時間領域の成分の前記第1の部分に含まれない部分からなる残りの部分を格納する重ね合わせ加算器、

をさらに備えている請求の範囲第2項に記載の装置。

7. 滑らかにされた時間領域の成分を生成して、前記雑音の低減された時間領域の成分の不連続性を最小化する後置ウィンドウ、および

前記滑らかにされた時間領域の成分の第1の部分を、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、前記滑らかにされた時間領域の成分の前記第1の部分に含まれない部分からなる残りの部分を格納する重ね合わせ加算器、

をさらに備えている請求の範囲第3項に記載の装置。

8. 雑音成分を有する入力音響信号に背景雑音の知覚される抑制を行うための装置であって、

前記入力音響信号から抽出される信号のフレームから周波数スペクトル成分を得る変換器、

各周波数成分に対して乗法利得因数を決定する検出器、

時間的分散およびスペクトル分散を生じさせるために、前記乗法利得因数を調整する分散機構、ならびに

前記周波数成分を変換して、雑音の修正されたスペクトル信号を取り出す制御減衰器、

を備えている装置。

9. 入力音響信号の背景雑音の知覚を低減するための方法であって、

前記入力音響信号を複数の信号フレームに分割し、

前記信号フレームから狭帯域の雑音成分を除去して、フィルタリングされた信号フレームを生成し、

一つのフィルタリングされた信号フレーム内の全信号を、前記一つのフィルタリングされた信号フレームに時間的に直前に先行するフィルタリングされた信号フレーム内のいくつかの信号と結合し、ウィンドウ化された信号フレームを生成し、

前記ウィンドウ化された信号フレームから周波数スペクトル成分を得、

前記周波数スペクトル成分を使用して、前記周波数スペクトル成分の雑音量の雑音評価値を生成し、

前記雑音評価値および前記周波数スペクトル成分に基づいて、利得乗法因数

を生成し、

前記周波数スペクトル成分を遅延させ、遅延した周波数スペクトル成分を生成し、

前記利得乗法因数に基づいて前記遅延した周波数スペクトル成分を減衰させ、雑音の低減された周波数成分を生成し、

前記雑音の低減された周波数成分を時間領域に変換する、
ステップを備えている方法。

10. 前記利得乗法因数を生成するステップが、

前記周波数スペクトル成分の各周波数成分に対して、その周波数成分が雑音であるかどうかについての決定を行い、

前記周波数スペクトル成分の各周波数成分に対して、周波数成分が雑音成分である確度レベルを導出し、

前記確度レベルに基づいて、各周波数成分に対して前記利得乗法因数を決定し、

前記利得乗法因数をスペクトルのおよび時間的に調整し、

前記周波数成分のスペクトルの谷間を検出および充填する、

ステップを備えている請求の範囲第9項に記載の方法。

11. 後置ウィンドウ化を行って、滑らかにされた時間領域の成分を生成し、

前記滑らかにされた時間領域の成分の第1の部分、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、

前記滑らかにされた時間領域の成分の前記第1の部分に含まれない部分からなる残りの部分を格納する、

ステップをさらに備えている請求の範囲第9項に記載の方法。

12. 後置ウィンドウ化を行って、滑らかにされた時間領域の成分を生成し、

前記滑らかにされた時間領域の成分の第1の部分、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、

前記滑らかにされた時間領域の成分の前記第1の部分に含まれない部分からなる残りの部分を格納する、

ステップをさらに備えている請求の範囲第10項に記載の方法。

13. 雑音成分を有する入力音響信号の背景雑音の知覚を低減するための方法であって、

前記入力音響信号から抽出される信号のフレームから周波数スペクトル成分を得、

各周波数成分に対して乗法利得因数を決定し、

時間的分散およびスペクトル分散を生じさせるために、前記乗法利得因数を調整し、

前記周波数成分を変換して、雑音の修正されたスペクトル信号を取り出す、ステップを備えている方法。

【発明の詳細な説明】**会話の品質向上のための背景雑音の低減****発明の背景**

本発明は、電話通信チャネルのようなチャネル上の音声情報の通信に関する。

音声送信システムに使用されるマイクロフォンは、典型的には、拾い集めるべき音声とともに、雑音と呼ばれる周囲の音、すなわち背景音を拾い集める。マイクロフォンから話し手まである距離がおかれている音声送信システム、例えば、映像音声電話会議環境で使用されるシステムにおいては、マイクロフォンによって拾い集められる会話に背景雑音が付加されるので、背景雑音は、音声品質の劣化の原因となる。それらの性質および使用目的によって、これらのシステムは、それらのマイクロフォンの周囲にある全ての位置からの音を拾い集めなければならず、これらの音は背景雑音を含んでいる。

HVACシステム、コンピュータ、および他の電子機器から発するファン雑音(fan noise)は、ほとんどの電話会議環境における支配的な雑音源である。

優れた雑音抑制技術は、背景雑音が知覚されることを低減する一方、同時に会話の質、すなわち会話の明瞭性に悪影響を与えない。一般に、本発明の目的は、一つのマイクロフォンによって拾われる会話に付加される狭帯域または広帯域の一定の雑音を抑制することにある。本発明の他の目的は、一つのマイクロフォンによって拾い集められる会話に付加されるファン雑音を低減することにある。

発明の要約

本発明の一特徴において、一般に、本発明は、入力音響信号の背景雑音を低減するための装置に関する。この装置は、入力音響信号を複数の信号フレームに分割するフレーム化装置、および前記信号フレームのそれぞれから雑音成分を除去して、フィルタリングされた信号フレームを生成するノッチ・フィルタ・バンクを特徴として備えている。乗算器は、結合された信号フレームを乗じ、ウィンドウ化された信号フレームを生成する。ここで、結合された信号フレームは、一つ

のフィルタリングされた信号フレームに時間的に直前に先行するフィルタリングされた信号フレーム内のいくつかの信号と結合された前記一つのフィルタリング

された信号フレーム内の全信号を含んでいる。変換器は、前記ウィンドウ化された信号フレームから周波数スペクトル成分を得、背景雑音評価器は、前記周波数スペクトル成分を使用して、前記周波数スペクトル成分の雑音量である雑音評価値を生成する。雑音抑制スペクトル修正器は、前記雑音評価値および前記周波数スペクトル成分に基づいて利得乗法因数を生成する。遅延器は、前記周波数スペクトル成分を遅延させ、遅延した周波数スペクトル成分を生成する。制御減衰器は、前記利得乗法因数に基づいて前記周波数スペクトル成分を減衰させ、雑音の低減された周波数成分を生成する。逆変換器は、前記雑音の低減された周波数成分を時間領域に変換する。

好ましい実施形態においては、前記雑音抑制スペクトル修正器は、グローバル判定機構、ローカル判定機構、検出器、分散機構およびスペクトル・バレー充填器を備えている。

グローバル判定機構は、前記周波数スペクトル成分の各周波数成分に対して、その周波数成分が主に雑音であるかどうかについての決定を行う。ローカル雑音判定機構は、前記周波数スペクトル成分の各周波数成分に対して、周波数成分が主として雑音成分である確度レベルを導出する。検出器は、前記確度レベルに基づいて、各周波数成分に対して前記利得乗法因数を決定する。分散機構は、前記決定された利得乗法因数の影響をスペクトル的および時間的に分散し、スペクトル・バレー充填器は、その結果の周波数成分のスペクトルの谷間を検出し、充填する。

また、この好ましい実施形態の変形例として、前記背景雑音評価器は、各周波数スペクトル成分に対して雑音評価値を生成し、前記ローカル雑音判定機構は、周波数成分のそれぞれとその対応する雑音評価値との比、および前記グローバル判定機構によって行われた決定に基づいて確度レベルを導出する。

他の実施形態において、本発明は、後置ウィンドウおよび重ね合わせ加算機構を特徴としてさらに備えている。後置ウィンドウは、滑らかにされた時間領域の成分を生成して、前記雑音の低減された時間領域の成分の不連続性を最小化する。

。

重ね合わせ加算機構は、前記滑らかにされた時間領域の成分の第1の部分、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、前記滑らかにされた周波数成分の前記第1の部分に含まれない部分からなる残りの部分を格納する。

本装置の好ましい実施形態において、背景雑音評価器は、それぞれが背景雑音評価値を生成する少なくとも二つの評価器、およびこれらの背景雑音評価値を比較し、その一つを選択する比較器を含んでいる。これらの評価器の一方は稼働最小評価器であり、他方の評価器は定常評価器である。

好ましい実施形態において、本装置は、ノッチ・フィルタ・バンクのためのノッチの位置を決定するノッチ・フィルタ機構も含んでいる。

図面の簡単な説明

図1は、本発明による雑音抑制システムのブロック図である。

図2～図4は、図1のブロック図の部分部分を実現する詳細なブロック図である。

好ましい実施形態の説明

世界中の何百万人もの人々によって日常使用されている最もシンプルな雑音抑制装置は、いわゆる「スケルチ」回路である。スケルチ回路は、ほとんどの市民バンド・2ウェイ・ラジオの標準である。この回路は、受信信号のエネルギーがあるしきい値より下がると、システムのラウドスピーカを単純に接続解除することにより作動する。このしきい値の値は、たいていは、マニュアルの制御ノブを使用して、遠端（far end）が静寂になるときに背景雑音がスピーカに渡らないようなレベルに固定される。この種の回路の問題は、遠端スピーカがスタートし、続いてストップするように回路がオン・オフすると、雑音が存在し、続いて消えることが明瞭に聞き取れるということである。雑音は広帯域で、小さな会話のエネルギーの存在する周波数をカバーするので、人が話しているのと同時に雑音が聞き取られる。どんなものでも雑音抑制の行われることが好ましいが、スケルチ・ユニットの動作は、非常に混乱した効果を生み出す。

本発明の雑音抑制方法は、音の会話と非会話の部分の双方における背景雑音を

低減することによって、「スケルチ」コンセプトを著しく優れたものとする。

本発明によるアプローチは、人間の知覚作用に基づいている。スペクトル・マスクおよび時間マスク（双方ともに以下に定義する）の原理を使用して、本発明は、会話信号に付加または混合される雑音の知覚される大きさを低減する。

このアプローチは、例えば、抑制システムの処理後の会話出力と会話成分のみ（無雑音会話）との間の平均自乗誤差を最小化することが目的である他のアプローチとは異なる。

本発明で使用される方法は、チャンネルのエネルギーがしきい値を超えると、そのチャンネルの利得を上げ、チャンネルのエネルギーがしきい値より下がると、利得を下げるという「スケルチ」の概念を利用するが、本方法は、異なる周波数領域において個別に処理を実行する。チャンネルの利得は、入力信号のボリュームとそれに対応する出力信号のボリュームとの比であるとみなすことができる。

さらに、本方法は、スペクトル・マスクを行うさまざまな心理音響学の原理を利用する。特に、本方法は、ある周波数において大きな音があると、その周波数の回りには、他の信号を聞き取ることができない所与の周波数帯域（臨界帯域と呼ばれる）が存在することを基本的に述べている原理を利用する。すなわち、臨界帯域にある他の信号は聞こえない。本発明の方法は、会話が遠端から受信されている間、雑音の知覚作用を低減させる簡単な「スケルチ」回路よりもはるかに効果的である。

また、本発明の方法は、一時的なマスクを行う特性を利用する。大きな音のバーストが発生すると、そのバースト後の200ミリ秒までの期間中、バーストのスペクトル領域において耳の感度が減少する。別の音響効果は、バースト前の20ミリ秒までの時間の間、耳の感度が減少するというものである（このように、人間のヒアリングは、約20ミリ秒のパイプライン遅延を有する）。したがって、本発明の一つのキーとなる要素は、十分大きな信号の発生前後の期間中、雑音に対する耳の感度が減少するので、所与の帯域の利得をそれ以下に下げる信号しきい値を、その帯域における十分大きな信号の発生前後において、一定期間小さくすることができるというものである。

システムの概要

図1のブロック図を参照して、入力信号1は、まず、フレーマ（フレーム化装置）2によって分割され、20ミリ秒のサンプル・フレームにされる（入力信号は16kHzのレートでサンプリングされるので、各20msのフレームは320個のサンプルを含む）。本方法の計算の複雑さは、一時に個々のサンプルに処理を行わずに、一時にサンプル・フレームのグループに処理を行うことにより大幅に減少する。続いて、フレームにされた信号は、ノッチ・フィルタ4のバンクに入力する。ノッチ・フィルタ4は、典型的にはモータの回転周波数で発生するモータ雑音のような狭帯域の雑音成分を除去するためのものである。ノッチが、十分疎らなスペクトル密度で十分狭いならば、会話の音質は悪影響を受けない。続いて、デジタル信号の各フレームは、直前のデジタル信号フレームの終端からの一部と組み合わせられ、ウィンドウ化フレームが生成される。

好ましい実施形態においては、デジタル信号の各フレーム（20ms）は、先行するフレームの最後の12msと結合されて、32msの間隔を有するウィンドウ化フレームにされる。つまり、各ウィンドウ化フレームは、あるデジタル信号フレームからの320個のサンプルと、直前のフレームのフィルタリングされたサンプルのうちの最後の192個のサンプルとを組み合わせで備えている。続いて、この512個のサンプルからなる会話セグメントは、乗算器6においてウィンドウを乗じられて、512サンプル・フレームの開始部と終端部において信号の不連続性から生じる問題が軽減される。続いて、高速フーリエ変換（FFT）8が、512個のサンプルからなるウィンドウ化フレームに施され、257個の成分の周波数スペクトルが生成される。

変換される信号の最低周波数（D.C.）成分と最高周波数（サンプリング周波数を2で割った値、すなわち8kHz）成分は、実数部のみを有し、他の255個の成分は、実数部と虚数部の双方を有する。このスペクトル成分は、背景雑音評価器20に与えられる。この評価器は、背景雑音のスペクトル・エネルギーを評価し、ノッチ・フィルタ4のノッチを配置すべき背景雑音のスペクトルのピークを見つけ出すためのものである。各周波数成分に対する信号強度のスペクトル評価器である定常評価器(stationary estimator)24および背景雑音のスペクトル評価器であ

る稼働最小評価器(running minimum estimator)22の評価値は、比較器28によって比較される。そして、特定の周波数成分が主に雑音からなるのか、信号と雑音の混ざったものからなるのかについての判定が、判定機構32により各周波数に対して行われることにより、さまざまな確度レベルが抽出される。これらの確度レベルに基づいて、周波数帯域の利得が、利得設定器34によって決定される。続いて、利得は、分散機構36によって、臨界帯域の周波数領域に分散され、スペクトル的にも時間的にも、心理音響学的なマスク効果が利用される。スペクトル・バレー充填器(spectral valley filler)38は、周波数成分利得関数におけるスペクトルの谷間(バレー:valley)を検出し、その谷間を埋めるために使用される。雑音圧縮スペクトル修正器30からの最終的な周波数成分利得関数は、減衰器12において、512ポイントFFTのスペクトル成分の大きさを修正するために使用される。減衰器12におけるフレームは、主として利得を生成するために使用される信号の後に続く1つの時間ユニットである。続いて、逆FFT(IFFT)14が、信号を周波数領域から時間領域に逆写像する。その結果として雑音が低減された信号の512ポイント・フレームは、乗算器16においてウィンドウを乗じられる。続いて、その結果は、加算器18において先行フレームの信号に重ね合わせおよび加算が施され、20ミリ秒、すなわち320個のサンプルの出力信号がライン40に出力される。

この信号処理の流れにおける各ブロックの詳細な説明は、入力から出力へその生起順に以下に示される。

上述したように、フレームにされた入力信号は、ノッチ・フィルタ4のバンクを介して引き出される。

図1および図2を参照して、ノッチ・フィルタ・バンク4は、無限インパルス応答(IIR: Infinite Impulse Response)デジタル・フィルタのカスケードからなる。各フィルタは、以下の式の応答を有する。

$$H(z) = \frac{1 - 2\cos\theta z^{-1} + z^{-2}}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \quad (1)$$

ここで、 $\theta = \pi/8000 \times$ (ノッチの周波数) であり、 r はノッチの幅を反映した1より小さな値である。 -3 dB幅のノッチが ω Hzであるならば、 $r = 1 - (\omega/2)$

($\pi/8000$)となる。例示する好ましい実施形態において使用される帯域幅 ω は、20Hzである。ノッチは、公称周波数に近い背景雑音エネルギーの最大ピークにおいて、約100 Hzごとに配置される。

ノッチ・フィルタリングは、新たな信号フレームの320個のサンプルに施される。フィルタリング結果の320個のサンプルは、先行するフレームからのノッチ・フィルタリングされた出力の最後の192個のサンプルに追加されて、全部で512個のサンプルからなる拡張フレームが生成される。

図1および図2を参照して、フィルタ・バンク4から取り出されたノッチ・フィルタリング後の512サンプル・フレームは、次の式を用いてウィンドウを乗じられる。

$$\begin{aligned} w(i) &= f(i) \sqrt{0.5 - 0.5 \cos\left(\pi \frac{i}{191}\right)} && \text{for } i=0, 1, \dots, 191 \\ w(i) &= f(i) && \text{for } i=192, 193, \dots, 319 \\ w(i) &= f(i) \sqrt{0.5 - 0.5 \cos\left(\pi \frac{511-i}{191}\right)} && \text{for } i=320, 321, \dots, 511 \end{aligned} \quad (2)$$

ここで、 $f(i)$ は、フィルタ・バンク4からの512サンプル・フレームのノッチ・フィルタリングされた第 i 番目のサンプルの値である。 $W(i)$ は、FFT8に次に与えられる512サンプルのウィンドウ化された出力結果の第 i 番目のサンプルの結果値である。乗算器6によってもたらされるウィンドウは、拡張フレームの開始部と終了部におけるエッジ効果および不連続性を最小化するためのものである。

時間ウィンドウ化された512個のサンプル・ポイントは、FFT8に与えられる。FFTの遍在性により、多くのデジタル信号処理(DSP)チップの製造者が、FFTを実現するための高度に最適化されたアセンブリ言語コードを供給している。

1フレーム・ディレイ10は、FFTの信号周波数成分を増幅でき、かつ、後に発生する信号値に基づいて減衰器12で処理できるように導入されている。これは

、知覚される雑音を導入しない。その理由は、信号成分は、上述したように、それが実際に発生する前のそのスペクトルの20ミリ秒の近傍部分の周波数をマスクするからである。また、会話音のボリュームは、零の振幅から開始してしだいに増大するので、この1フレームのディレイは、会話の発声開始のクリッピングを防止する。

雑音によるFFTの成分は、減衰器12によって減衰するが、一方、信号による成分の減衰量は少ないか、もしくは減衰せず、または増幅されることもある。上述したように、各周波数に対して、実数成分と虚数成分がある。両成分は、雑音抑制スペクトル修正モジュール30から見つけ出された単一の因数(factor)を乗じられ、位相は、周波数成分の大きさが変更されても、その周波数成分に対して維持される。

逆FFT14(IFFT)は、大きさの修正されたFFTに施され、長さが512個のサンプルからなる周波数処理された拡張フレームが生成される。

乗算器16で使用されるウィンドウ化処理は、上述した乗算器6のウィンドウ化処理と全く同じである。その目的は、周波数成分の減衰により発生する不連続性を最小化することである。例えば、全ての周波数成分が1つを除いて零に設定されていると仮定する。IFFTを施した結果は正弦波になる。この正弦波は、大きな値で始まり、大きな値で終わることがある。隣接するフレームには、この正弦波成分が存在しないことがある。したがって、適切なウィンドウ化処理を行うことなく、出力加算器18でこの信号に重ね合わせおよび加算が行われると、クリック音がフレームの開始部および終了部で聞こえることがある。しかしながら、例えば、式2で定めたパラメータを使用して、正弦波を適切にウィンドウ化することによって、大きさがスムーズに増加しスムーズに減少する正弦波が聞こえることになる。

乗算器6および16によるフレームの前置ウィンドウ化処理および後置ウィンドウ化処理のために、フレームの重ね合わせおよび加算処理は、フレームの開始部および終了部において出力の大きさが減少することを防止するために必要となる。

したがって、512個のサンプルからなる現在の拡張ウィンドウ化フレームの最初の192個のサンプルは、先行の拡張ウィンドウ化フレームの最後の192個のサンプルに加えられる。続いて、現在の拡張フレームの次の128個のサンプル（8ミリ秒）が出力される。続いて、現在の拡張ウィンドウ化フレームの最後の192個のサンプルは、次のフレームとの重ね合わせ／加算処理等に使用するために格納される。

好ましい実施形態において、使用されるウィンドウ関数 W は、変調時間の超過を避けるために、次の特性を有する。

$$W^2 + (\text{重ね合わせ量だけシフトした} W^2) = 1$$

例えば、重ね合わせ量がフレームの $1/2$ ならば、ウィンドウ関数 W は次の特性を有する。

$$W^2 + (1/2 \text{ だけシフトした} W^2) = 1$$

[背景雑音評価器20]

図1および図3を参照して、背景雑音評価器20および雑音抑制スペクトル修正モジュール30は、次のように動作する。

背景雑音評価器20は、FFTの各周波数成分の評価を行い、背景雑音によるエネルギーの大きさの平均値を求めるためのものである。この背景雑音評価器により、環境が変わるごとに、ユーザがシステムをマニュアルで調整または操作する必要はなくなる。背景雑音評価器は、信号／雑音環境を絶えず監視し、例えば、エアコンのファンがオン・オフする等に応答して、自動的に背景雑音の評価値を更新する。2つのアプローチが、ある特定の状況で使用される一方のアプローチまたは他方のアプローチの結果とともに使用される。第1のアプローチは、より正確ではあるが、背景雑音のみに対して1秒間隔を必要とする。第2のアプローチは、正確さでは劣るが、どのような状況であっても、10秒で背景雑音評価値を求める。

[定常評価器24]

図1および図3を参照して、第1のアプローチは、定常評価器24を使用して、各フレームのスペクトル形状が他のフレームのスペクトル形状と非常に類似する長いフレーム・シーケンスを探す。たぶん、このような状況は、部屋にいる人間

は静かにしており、ファンや回路の雑音による一定の背景雑音が主な信号源である場合にのみ生じ得る。このようなシーケンスが検出されると、各周波数の平均の大きさは、F F Tシーケンスの中央部分のフレームから取られる（シーケンスの開始部および終了部のフレームは、低レベルの会話成分を含んでいることがある）。この方法は、第2のアプローチ（後述）よりもはるかに正確に背景雑音のスペクトルを測定するが、背景雑音が比較的一定であることと、部屋にいる人間がある一定期間会話をしていないことを必要とし、このような状況は、実際にはあまり見られない状況である。

この評価器の詳細な処理は、次に示すとおりである。

1. 図3を参照して、第1のアプローチによる方法は、現在の20msフレームのスペクトル形状が先行フレームのスペクトル形状と類似しているかどうかを判断する。まず、この方法は、ステップ240において、先行フレームのスペクトル形状を計算する。

$$N_1(f_c) = 0.25 \sum_{f=f_c-3}^{f=f_c+3} \left(\sum_{k=k_i}^{k=k_i+31} (R^2(k, f) + I^2(k, f)) \right) \quad (3)$$

ここで、 f_c は、現在の20msフレームのフレーム番号である（この番号は、連続フレームの一つごとに進められる）。 i は、1000Hzの周波数帯域を示し、 $k_i = i \times 32$ である。 k は、512ポイントF F Tの256個の周波数成分のインデックスである。 $R(k, f)$ と $I(k, f)$ は、それぞれ、フレーム f の第 k 番目の周波数成分の実数成分と虚数成分である。

2. 次に、現在のフレームのスペクトル形状 $S_i(f_c)$ が、ステップ242で決定される。

$$S_i(f_c) = \sum_{k=k_i}^{k=k_i+31} (R^2(k, f_c) + I^2(k, f_c)) \quad (4)$$

ここで、この式は、上記式(3)と同じ意味を有する。 S_i は、現在のフレーム f_c の第 i 番目の周波数成分の大きさである。

3. 続いて、評価器24は、ステップ244および246において、以下の不等式が成立するかどうかをチェックする。

$$N_i(f_c) > t_l S_i(f_c) \quad (5)$$

または

$$S_i(f_c) > t_l N_i(f_c), \text{ for } i = 0, 1, \dots, 7 \quad (6)$$

ここで、 t_l は低い方のしきい値である。好ましい実施形態においては、 $t_l = 3$ である。不等式(5)または(6)が、 i の4つの値よりも多い値に対して成立するならば、現在のフレーム f_c は、信号として分類される。そうでなければ、評価器は、次の不等式が成立するかどうかをチェックする。

$$N_i(f_c) > t_h S_i(f_c) \quad (7)$$

または

$$S_i(f_c) > t_h N_i(f_c), \text{ for } i = 0, 1, \dots, 7 \quad (8)$$

ここで、 t_h は高い方のしきい値である。 N_i は、背景雑音評価値の第 i 番目の周波数成分の大きさを示す。好ましい実施形態においては、 $t_h = 4.5$ である。 i の1つまたは2つ以上の値に対して、いずれかの不等式が成立すると、現在のフレーム f_c は信号フレームとして分類される。そうでなければ、現在のフレームは、雑音として分類される。

4. ステップ252において、雑音に分類されたフレームが50個連続して発生すると(1秒間の雑音に相当する)、評価器24は、第10番目から第41番目のフレームの周波数エネルギーを合計することにより、背景雑音の評価値を求める。このシーケンスの開始部分のフレームおよび終了部分のフレームを無視することにより、信号が残りのフレームに存在しないという確度が増加する。評価器は、ステップ254において、次の式の値を求める。

$$B_k = \frac{1}{32} \sum_{f=f_s}^{f=f_s+31} (R^2(k, f) + I^2(k, f)) \quad (9)$$

ここで、 $k = 0, 1, 2, \dots, 255$ であり、 f_s は、第10番目の雑音に分類されたフレームの開始インデックスである。他の変数は、式(3)と同一の表記を有する。値 B_k は、第 k 番目の周波数に対する信号の雑音成分スペクトルの平均の

大きさを表す。

図1および図4を参照して、ノッチ・フィルタ・バンクのノッチを配置する場所を決定するために、32個の雑音に分類されたフレームのみに対応するウィンドウ化されていない20msの時間領域サンプルが互いに結合され、連続シーケンスを形成する（ステップ260）。このシーケンスに対して、長いFFTが施される（ステップ262）。約100 Hzごとに最大の大きさを有する成分が、突き止められる（ステップ264）。この局所的に最大の大きさが生じる周波数が、ノッチの中心周波数が置かれる位置に対応する（ステップ266）。ノッチは、1500Hz辺りまでのファン雑音のみを減少させるのに有益である。その理由は、ファン雑音スペクトルは、より高い周波数に対して強いピークが存在せず、かなり平坦になる傾向を有するからである。

〔稼働最小評価器22〕

会話信号が1秒間以上存在しないということがないか、背景雑音自体が一定のスペクトル形状でない場合がある。この場合には、定常評価器24（上述）は、背景雑音評価値を生成しない。このような場合のために、稼働最小評価器22が、正確さの点では劣るものの、背景雑音評価値を生成する。

稼働最小評価器によって使用される処理は、次のとおりである。

1. 10秒間隔に渡って、各周波数成分 k に対して、その周波数成分についての8つの連続フレームのエネルギーを最小にする8つの連続フレームを見つけ出す。すなわち、すべての周波数成分 k に対して、以下の $M_k(f_k)$ を最小にするフレーム f_k を見つけ出す。

$$M_k(f_k) = \frac{1}{8} \sum_{f=f_k}^{f=f_k+7} (R^2(k, f) + I^2(k, f)) \quad (10)$$

ここで、 f_k は、上記10秒間隔内で発生する任意のフレーム番号である。一般に、式(10)を最小にする f_k は、異なる周波数成分 k に対して異なる値をとる。

2. 以下の2つの状況が両方とも成立すると、先のステップで背景雑音スペクトル評価値として求められた最小値 M_k を使用する。

(a) 定常評価器による背景雑音スペクトル評価値の最後の更新から10秒以上が

経過している。

(b) 過去の背景雑音評価値（定常評価器または稼働最小評価器から求められたもの）と現在の稼働最小評価器との差Dが大きい。差Dを定義するために使用される基準は、式(11)で与えられる。

$$D = \sum_{k=0}^{K-255} \left(\max \left(\frac{M_k}{N_k}, \frac{N_k}{M_k} \right) - 1 \right)^2 \quad (11)$$

ここで、max関数は、その2つの引数の大きい方の値を返す。N_kは、（稼働最小評価器または定常最小評価器のいずれかからの）前の背景雑音評価値である。M_kは、稼働最小評価器からの現在の背景雑音評価値である。

Dが、あるしきい値（例えば、好ましい実施形態においては3000）よりも大きく、上記条件(a)を満足しているならば、M_kは、新しい背景スペクトル評価値として使用される。M_kを雑音評価値として使用することは、ノッチ・フィルタを不作動にすべきことを示す。その理由は、ノッチの中心周波数の優れた評価が可能でないからである。

[雑音抑制スペクトル修正器30]

図1を参照して、背景雑音評価値が求められると、現在のフレームのスペクトルを背景雑音評価値のスペクトルと比較し、この比較に基づいて、出力信号の雑音の知覚を減少させるために、現在のフレームのFFTの各周波数成分に対して減衰処理を行わなければならない。

[グローバル会話信号対雑音検出器32]

任意の所与のフレームは、会話信号とそれ以外のもののいずれかを含んでいる。グローバル会話信号対雑音検出器(Global speech versus noise detector)32は、フレームが雑音であるかどうかについて二値の判定を行う。

会話信号が存在する場合に、マスク効果が不正確な信号対雑音宣言を目立たなくさせる傾向にあるので、しきい値を低くすることができる。一方、フレームが雑音のみであるならば、周波数成分が雑音によるものか信号によるものを決定することの僅かな誤りが、いわゆる「きらめき」音(twinkling sound)を引き起こすことになる。

例示の実施形態によると、フレーム内に会話信号が存在するかどうかを判断するために、システムは、現在のフレームの第 k 番目の周波数成分の大きさ S_k と背景雑音評価値の第 k 番目の周波数成分の大きさ C_k とを比較する。次に、(1 フレームに対して) k の 7 つの値より多い値について $S_k > T \times C_k$ ならば、そのフレームは会話信号のフレームと宣言される。そうでなければ、そのフレームは雑音フレームと宣言される。ここで、 T は一定のしきい値である (好ましい実施形態では $T = 3$)。

[個別周波数成分用ローカル会話信号対雑音検出器34]

上記節で説明したグローバル会話信号対雑音検出器32は、各周波数成分が雑音かどうかについて二値判定を行うものである。一方、ローカル会話信号対雑音検出器(Local speech versus noise detector)34は、各周波数成分が雑音であるかどうかについて幅のある判定を行う。これらの判定は、第 k 番目の周波数成分が雑音であるという高い確度の判定から、第 k 番目の周波数成分が信号であるという高い確度の判定までの幅を有する。

この判定は、現在のフレームの第 k 番目周波数成分の大きさの、背景雑音スペクトル評価値のうちのこれに対応する成分の大きさに対する比に基づいてなされる。判定値を D_k で示す。この実施形態において、判定値 D_k は 0 から 4 まで変動し、 $D_k = 0$ の判定は、「その成分が雑音である確度が高い」ことに対応し、 $D_k = 4$ の判定は、「その成分が信号である確度が高い」ことに対応する。

D_k の値は以下のように定められる。

$$\begin{aligned}
 & \text{if } \frac{S_k}{N_k} > t_4, D_k = 4, \\
 & \text{else if } \frac{S_k}{N_k} > t_3, D_k = 3, \\
 & \text{else if } \frac{S_k}{N_k} > t_2, D_k = 2, \\
 & \text{else if } \frac{S_k}{N_k} > t_1, D_k = 1, \\
 & \text{else } D_k = 0
 \end{aligned} \tag{12}$$

ここで、現在のフレームに対して、 $S_k = R^2(k) + I^2(k)$ である。 N_k は成分 k の背景雑音評価値である。 t_1 、 t_2 、 t_3 、 t_4 に使用される値は、グローバル会話信号検出器32が先行フレームを会話信号と宣言したか雑音と宣言したかに

依存して変化する。例示の実施形態において、雑音と宣言した場合には、 $t_1=6.3$ 、 $t_2=9.46$ 、 $t_3=18.9$ 、 $t_4=25.2$ であり、信号と宣言した場合には、これらのしきい値は $1/2$ に下げられ、 $t_1=3.15$ 、 $t_2=4.73$ 、 $t_3=9.45$ 、 $t_4=12.6$ にされる。

これらの D_k は、制御される減衰器の利得乗法因数(gain multiplicative factor)を調整する以下の処理に使用される。

[臨界帯域付近の周波数ビン(frequency bin)利得の時間スペクトル分散器36]

配列 A_k は、すべてのFFT周波数成分 k に対する乗法因数(multiplicative factor)を格納する。 A_k の要素は、FFT8から1フレーム・ディレイ10を介して与えられるスペクトル成分を修正するために、制御減衰器12によって使用される。 A_k の値は、ローカル会話信号対雑音検出器34で求められた判定値 D_k に基づいて変更される。

A_k の値は、 $L < A_k < 1$ の範囲に制限される。ここで、 L は、雑音低減量の下限である(以下に説明する)。 L の値が小さいほど、より多くの雑音を低減することができる。しかしながら、一般に、雑音を多く低減するほど、人為的なもの(artifact)が同時に発生してくる。信号の信号対雑音比(SN比)を高くすると、反対すべき人為的なものを会話信号内に生成することなく、 L の値を小さく設定することができる。適度な14dBのSN比に対する L の良好な値は、0.25である。会話の明瞭性に影響する反対すべき人為物を減らすために、SN比が低くなるにつれて、 L の値を大きくすべきである。例えば、6dBのSN比では、 L の値を0.5にする必要がある。会話のSN比が、システムの処理中に測定され、測定値は、 L の値を決定するために使用される。

A_k は、上記式(12)から導出されるように、先行フレームにおける A_k と現在のフレームの D_k の値の関数として、各新フレームごとに変化する。第 i 番目のフレームからの A_k を $A_{k,i}$ とすると、 $A_{k,i} = G(A_{k,i-1}, D_k)$ となる。ここで、関数 G は、以下の式(13)で定義される。

$$\begin{aligned}
 G(A_k^i, D_k) &= \text{if } D_k = 0 \text{ then} \\
 &\quad A_k^i = A_k^{i-1} \times \beta_0 \\
 &\quad \text{if } D_k \geq 1 \text{ then} \\
 &\quad \quad A_k^i = A_k^{i-1} \times \beta
 \end{aligned}
 \tag{13}$$

ここで、 $\beta > 1$ であり、 D_k とともに増加する。 $\beta_0 < 1$ である。

つまり、判定値 $D_k \geq 1$ ならば、先行フレームからの A_k は、 D_k の値の増加とともに増加する 1 より大きな乗法因数を乗じられる。判定値 $D_k = 0$ ならば、先行フレームからの A_k は、1 より小さな値（代表的には 0.8）の乗法因数を乗じられる。

これは、時間的分散（拡張、拡散：temporal spreading）である。好ましい実施形態では、時間的分散は、現在のフレーム前 20ms から現在のフレーム後 200ms まで存在する。

判定値 $D_k = 4$ は、スペクトル成分 k が会話信号を含む確度が高いことを意味し、 A_k はその最大値 1 に設定される。

次に、 A_k のスペクトル分散（拡張、拡散：spectral spreading）が 500 Hz よりも大きな周波数に対して実行される。この分散は、この例示の実施形態においては中心周波数の 1/6 に等しい臨界帯域幅上で起こる。この理由は、心理音響学の実験からきている。この実験は、所与の周波数において強い信号成分があると、この信号成分は、その周波数の 1/6 の帯域幅における雑音に対してマスク効果を有する、ということを示している。スペクトル分散は、次のようにして達成される。

$D_k \geq 1$ の判定値に対して、上述したように計算される臨界帯域幅の A_k は、 $F(D_k)$ より小さくなることは許されない（ F は、以下で定義する）。本質的に、雑音（および会話信号）の低減量がスペクトル領域で少なくなるように、臨界帯域幅における A_k に対する下限 L は大きくされる。典型的には、 L は 0.25 に等しく、 $F(D_k)$ は、次のように定められる。

$$F(4) = 0.5$$

$$F(3) = 0.4$$

$$F(2) = 0.333$$

500 Hzより小さな周波数に対して、スペクトル分散は実行されない。実験では、背景雑音スペクトルは、低い周波数において、回転音響効果により、多くのピークと谷間を有し、耳障りな粗い人為的なものは、これらの周波数においてスペクトル分散を行うことが原因で生じることが示されている。

[スペクトル・バレー充填器38]

雑音低減方法の人為的なものの一つとして、残響感の増加がある。これは、残響による信号のスペクトル消失点(spectral nulls)の深さが、利得乗法因数の配列 A_k を導出するプロセスにおいて増加するという事実から起因する。この人為的なものとの戦いを補助するために、残響に伴うスペクトル消失点(null)に対応する A_k の局所的最小値が増加させられる。500 Hzから上の周波数の範囲に対して、 $A_k < A_{k-4}$ かつ $A_k < A_{k+4}$ であるならば、 $k = 16, 17, 18, \dots, 251$ に対して、次のようになる。

$$A_k = \frac{A_{k-4} + A_{k+4}}{2} \quad (14)$$

[減衰器12]

乗法因数配列の値がある特定のフレームについて決定されると、雑音対会話信号の決定に反映するように周波数成分を調整することができる。

制御される減衰器12において、乗法因数配列 A_k は次のように使用される。遅延変換された信号の実数および虚数の各周波数成分は、次のように増減される。

$$\begin{aligned} \text{for } k &= 0, 1, 2, \dots, 256 \\ R_n(k) &= A_k R(k) \\ I_n(k) &= A_k I(k) \end{aligned} \quad (15)$$

続いて、図1のブロック図に示すように、修正されたフーリエ成分 $R_n(k)$ および $I_n(k)$ には、逆FFT、ウィンドウ化および重ね合わせが行われ、最終的に、雑音が低減された出力信号フレームが生成される。

これにより、審美的に満足され、雑音が低減されて知覚されない信号が生成される。

本発明の好ましい特定の実施形態に対して付加し、減らし、削除する修正、および他の修正を行うことは、この技術分野の専門家（当業者）には明らかであり、以下の請求の範囲に含まれるものである。

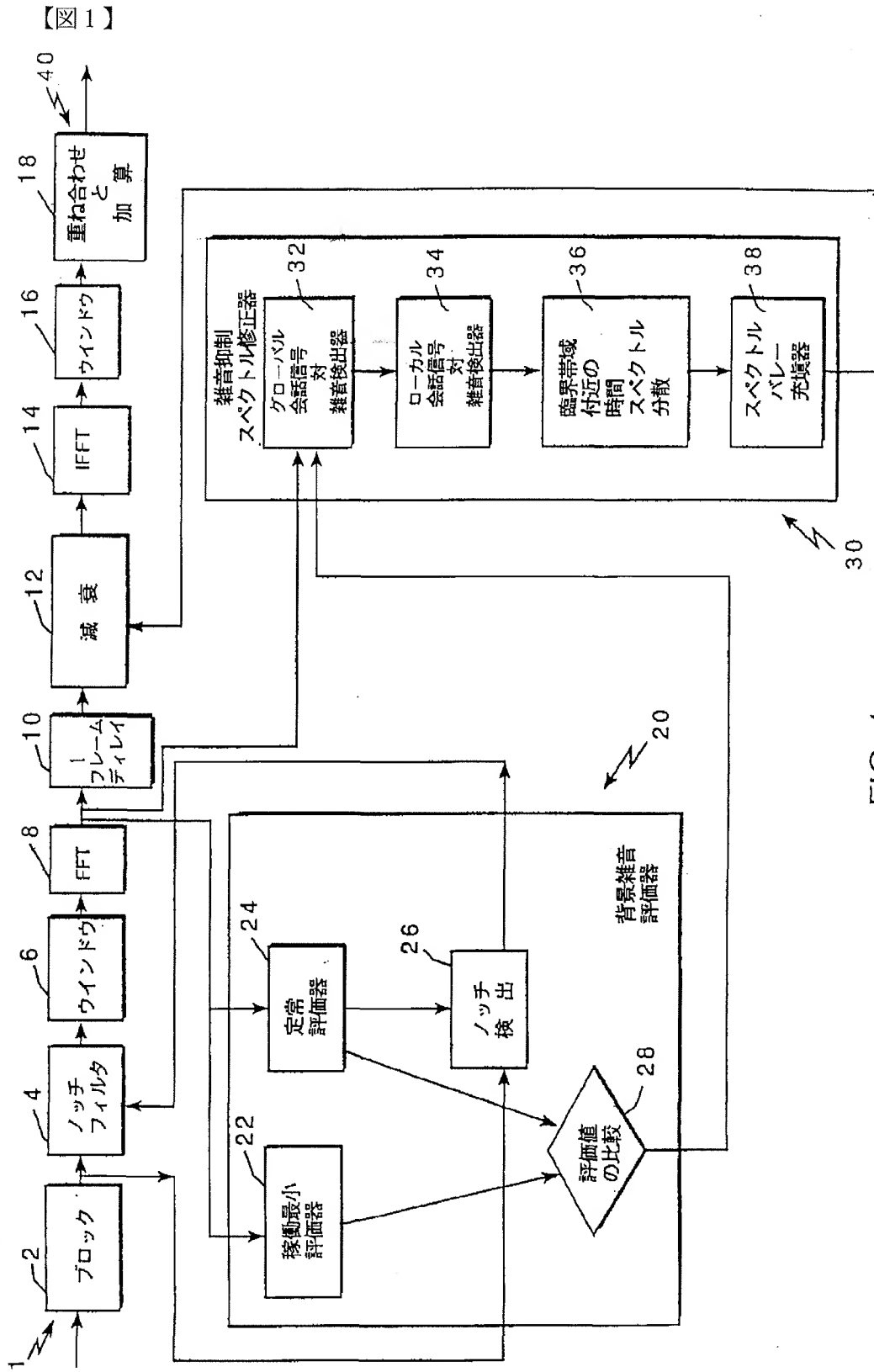


FIG. 1

【図2】

ノッチ・フィルタ

$$H(z) = \frac{1 - 2 \cos \theta z^{-1} + z^{-2}}{1 - 2r \cos \theta z^{-1} + r^2 z^{-2}} \quad (4)$$

ウインドウ

$$\begin{aligned} w(i) &= f(i) \sqrt{0.5 - 0.5 \cos \left(\pi \frac{i}{191} \right)} & \text{for } i = 0, 1, \dots, 191, \\ w(i) &= f(i) & \text{for } i = 192, 193, \dots, 319, \\ w(i) &= f(i) \sqrt{0.5 - 0.5 \cos \left(\pi \frac{511 - i}{191} \right)} & \text{for } i = 319, 320, \dots, 511 \end{aligned} \quad (6)$$

FIG. 2

【図3】

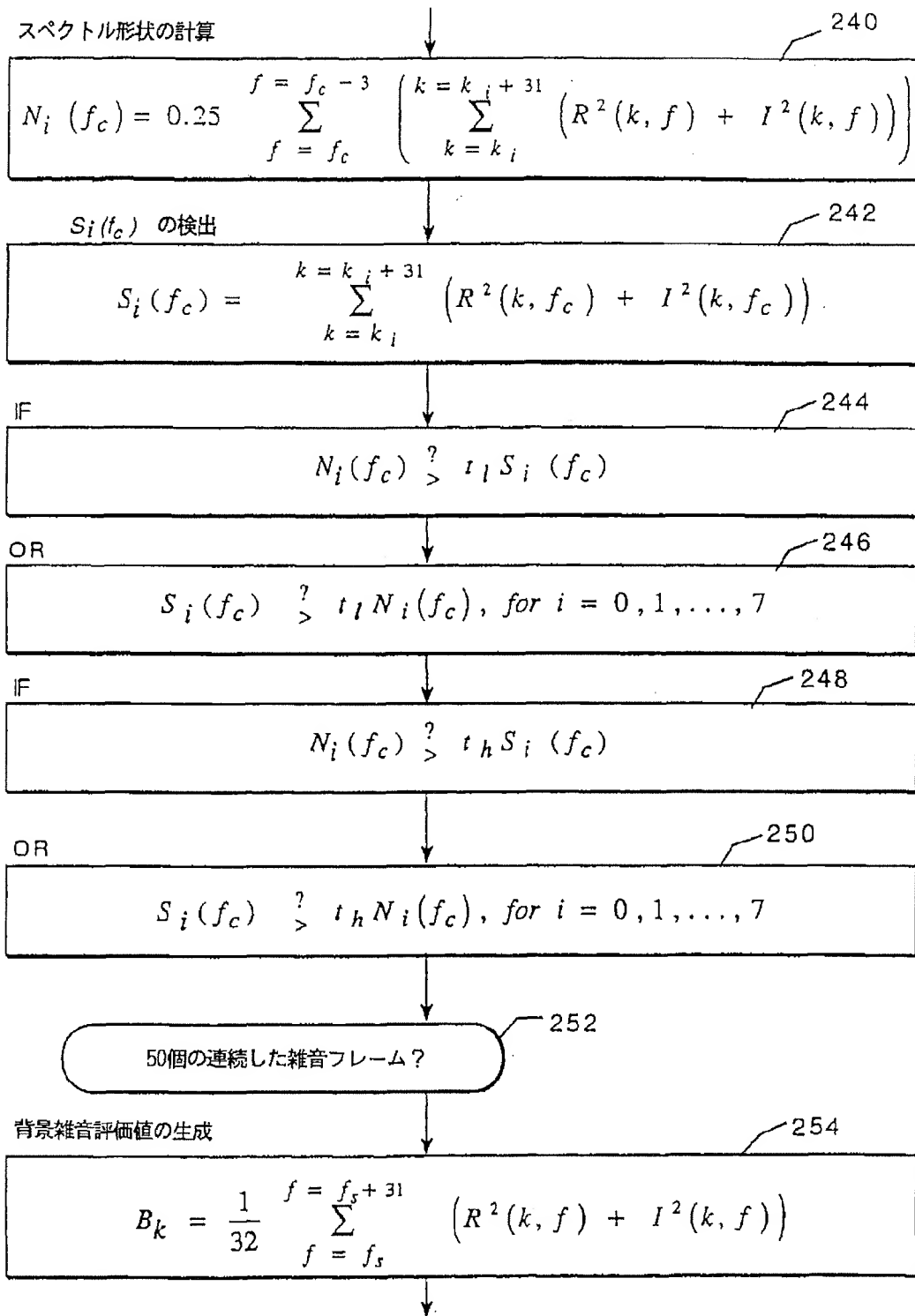


FIG. 3

【図4】

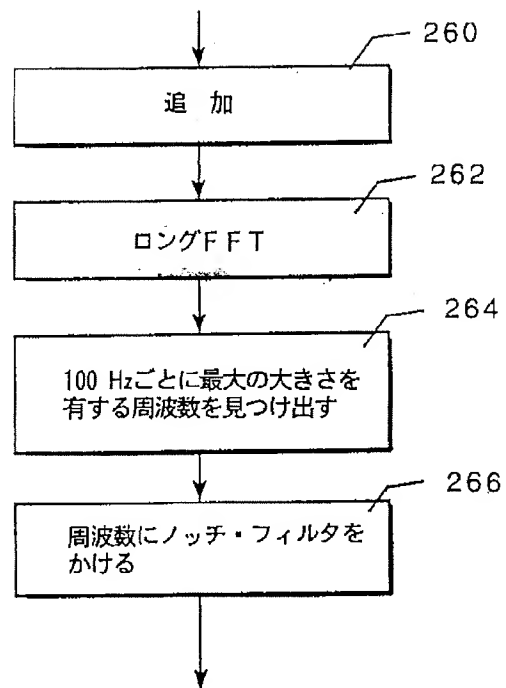


FIG. 4

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/06367

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : H04B 15/00

US CL : 381/94

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 381/94, 46, 47

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US,A 4,696,039 (DODDINGTON) 22 SEPTEMBER 1987, COLUMNS 1, 2 AND 3.	1,5-9,11,13
Y	US,A 5,012,519 (ADLERSBERG, ET AL.) 30 APRIL 1991, COLUMNS 4, 5 AND 6, FIGURES 1-4.	1, 5-9, 11, 13

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to underpin the principle or theory underlying the invention
"A" documents defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combinations being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"a"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

16 AUGUST 1994

Date of mailing of the international search report

OCT 27 1994

 Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

 Authorized officer

 MARK KELLY

Telephone No. (703) 305-4700

Form PCT/ISA/210 (second sheet)(July 1992)w